Relational Scaffolding Enhances Children's Understanding of Scientific Models

Benjamin D. Jee[1] and Florencia K. Anggoro[2]

[1]Worcester State University, Department of Psychology

[2]College of the Holy Cross, Department of Psychology

Author Note

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A150228 to the College of the Holy Cross. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. David Uttal, Cherilynn Morrow, and Dedre Gentner provided valuable feedback throughout the project. We thank the project coordinators and research assistants from College of the Holy Cross and Worcester State University who assisted with data collection and coding. We are grateful to the schools, teachers, parents, and students who took part in the research.

Correspondence concerning this article should be addressed to Benjamin Jee, Department of Psychology, Worcester State University. Contact: bjee@worcester.edu.

Abstract

Models are central to the practice and teaching of science. Yet people often fail to grasp how scientific models explain their observations of the world. Realizing the explanatory power of a model may require aligning its relational structure to that of the observable phenomena. The present research tested whether *relational scaffolding*—guided comparisons between observable and modeled events—enhances children's understanding of scientific models. We tested relational scaffolding during instruction about the day-night cycle, a topic that involves relating *Earth-based* observations to a *space-based* model of Earth rotation. Experiment 1 found that 3rd graders ($N = 108$) learned more from instruction that incorporated relational scaffolding. Experiment 2 ($N = 99$) found that guided comparison—not merely viewing observable and modeled events—is a critical component of relational scaffolding, especially for children with low initial knowledge. Relational scaffolding could be applied broadly to assist the many students who struggle with science.

Keywords: relational learning, structural alignment, comparison, relational scaffolding, science education, model-based instruction

Relational Scaffolding Enhances Children's Understanding of Scientific Models

Science portrays a world governed by invisible entities and processes. The orbiting of electrons around the nucleus of an atom or of planets around the Sun in our solar system cannot be directly observed. One is imperceptibly small, the other incredibly vast. The existence of these invisible systems is hardly common sense. Indeed, revolutionary scientific breakthroughs—such as the heliocentric model of the solar system, evolution by natural selection, and the germ theory of disease—have drastically restructured our understanding of the world (Kuhn, 1962). The advancement of these and other fundamental scientific ideas has involved creating, testing, and refining models that represent the hidden nature of reality (Nersessian, 2010).

Science education aims to support model-based learning from a young age (American Association for the Advancement of Science (AAAS), 2009; National Research Council (NRC), 2012). Yet, students often emerge from school with incomplete and incorrect ideas about the way the world works (Chi, Roscoe, Slotta, Roy, & Chase, 2012; McCloskey, 1983; Shtulman, 2017). To realize the explanatory power of a scientific model, a student must grasp its relation to observable phenomena. This is straightforward when models depict familiar objects in idealized form, or events happening more slowly or quickly than normal. It is far more challenging when models portray invisible entities and processes, whose relationship to observable phenomena is nonobvious and often counterintuitive.

Like interpreting an analogy, making sense of a scientific model can involve relating seemingly disparate sets of information. Hence, also like analogy, the mapping between observations and model may depend on a process of *structural alignment* in which correspondences are established in accordance with a deeper, shared system of relations (Gentner

& Markman, 1997). With scientific models, causal attribution is also crucial; the model *explains* what is observed. Considering these parallels, theory and research on analogical thinking could provide a basis for new approaches to model-based science instruction.

Our aim was to test a support for model-based learning—*relational scaffolding*—that involves guiding a student through systematic comparisons between observable phenomena and corresponding modeled events. Comparing analogous cases brings common relational structure into focus (Gentner & Markman, 1997). When cases lack surface similarity, explicit comparison supports analogical retrieval and mapping (Goldwater & Gentner, 2015; Holyoak & Koh, 1987; Kurtz, Miao, Gentner, 2001). Thus, relational scaffolding should be most effective when models involve entities and processes that bear little resemblance to observable phenomena.

Applying relational scaffolding throughout a complex system of relations could illuminate the system's structural coherence—its *systematicity*—and discourage students from fragmented, incoherent explanations (Gentner & Toupin, 1986). Comprehensive visual comparisons also reduce the burden of mentally representing and aligning observed and modeled events during instruction (Mayer & Moreno, 2003; Richland, Zur, & Holyoak, 2007). Thus, relational scaffolding should especially benefit students with little prior knowledge, who often misinterpret scientific explanations and experience cognitive overload during instruction (Kalyuga, 2007; McNamara, Kintsch, Songer, & Kintsch, 1996; Vosniadou & Skopeliti, 2017).

The present research tested relational scaffolding with a fundamental and notoriously challenging topic: the day-night cycle. Children in the US are expected to understand this topic between 3rd and 5th grade (NRC, 2012). However, many 3rd- and 5th-graders are confused about the connection between Earth motion and the day-night cycle, e.g., stating that *the Earth orbits the Sun in a single day* (Vosniadou & Brewer, 1994). Even 7th and 8th grade US students frequently endorse incorrect explanations (Sadler, Coyle, Miller, Cook-Smith, Dussault, & Gould, 2010).

Grasping the scientific model of the day-night cycle involves reconciling observations from an *Earth-based* perspective—including sunrise, midday, sunset, and midnight—with models that adopt a large-scale *space-based* perspective of the Earth-Sun system (see Figure 1). Relational scaffolding supports structural alignment by showing video footage of these two perspectives *simultaneously* (in a split-screen display) as an instructor indicates the relevant temporal, spatial, and causal correspondences.

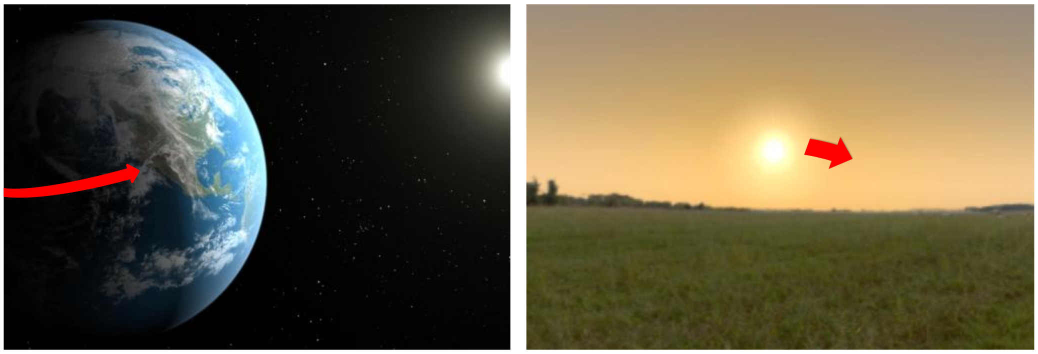


Figure 1. Earth's rotation from a space-based perspective (left), and the apparent motion of the Sun from an Earth-based perspective (right). Sources: nasa.gov (left), Stellarium.org (right). We added the arrows.

Relational scaffolding is intended to supplement rather than replace model-based science instruction. We therefore tested it in a sequence of instruction that progressed from a familiar frame of reference—an embodied simulation in which the student enacted Earth rotation—to a space-based simulation using a physical 3D model. Embodied simulation was intended to provide an intuitive, bridging analog for the external model (Clement, 1993), and to clarify the alignment between apparent and actual motion through direct physical experience (Kontra, Lyons, Fischer, & Beilock, 2015). Relational scaffolding incorporated video footage of each simulation activity, recorded from a 3rd-person/space-based perspective and a 1st-person/Earth-based perspective, as shown in Figure 2.

We tested the effects of relational scaffolding in two experiments. In Experiment 1, 3rd-grade participants were randomly assigned to one of four conditions. One group completed the sequence of instruction outlined above, progressing from an embodied to a 3D model simulation activity with supplementary relational scaffolding (RS condition). A second group completed the same two simulations but without relational scaffolding (No RS condition). A third repeated the 3D model simulation several times (Models Only condition). A fourth received no instruction (Control). We interviewed students about the day-night cycle before and after instruction. If relational scaffolding helps students relate their observations to a scientific model, participants in the RS condition should acquire the most knowledge. Moreover, if relational scaffolding is especially effective for students with little prior knowledge, any advantage of the RS condition should be greater for low-knowledge participants.

To assess changes in broader domain knowledge, we included at pre- and posttest items from a space science concept inventory (Sadler et al., 2010). Given the role of language in conceptual learning (e.g., Gentner, Anggoro, & Klibanoff, 2011) and of spatial thinking in space science understanding (e.g., Plummer, Kocareli, & Slagle, 2014), we assessed verbal and spatial abilities prior to instruction and included these factors in the analyses of student learning outcomes. Spatial tests were also administered at delayed posttest to explore potential changes resulting from the spatially-intense instruction.

Experiment 1

Method

Participants. One hundred forty-seven 3rd-grade children were sampled from public elementary schools in Worcester, Massachusetts. Thirty-six (24%) left the study before the session 6 posttest (attrition is discussed further in the Supplemental Online Materials). Three participants were removed from the dataset for scoring more than 2 *SD* below age level on a vocabulary assessment (the PPVT; see below). The remaining sample consisted of 108 3rd-graders (62 females, 46 males; M_{age} = 8.6 years, *SD* = 0.5). This sample size provided > .90 power to detect a medium-to-large effect size (Cohen's $f^2 \geq .25$) for our planned linear multiple regression analysis (G*Power; Faul, Erdfelder, Buchner, & Lang, 2009). Participants were randomly assigned to one of the four conditions (27 per condition).

Knowledge Measures

Day-Night Cycle Interview. This was our primary measure of understanding. Interviewers followed a script of questions and follow-ups (see Pretest Interview Script in the Supplemental Online Materials). Several questions required a verbal response. Others involved the use of 3D objects (rubber balls representing "Earth" and "Sun") to model specific events in the day-night cycle.

Items from the Astronomy and Space Science Concept Inventory (ASSCI). The ASSCI measures students' mastery of astronomical concepts found in national standards (Sadler et al., 2010). We selected nine items (seven K-4 and two 5-8 level) broadly related to, but not covered in, our instruction. The Supplementary Online Materials provide further detail on the instrument and selected items.

Cognitive Ability Measures

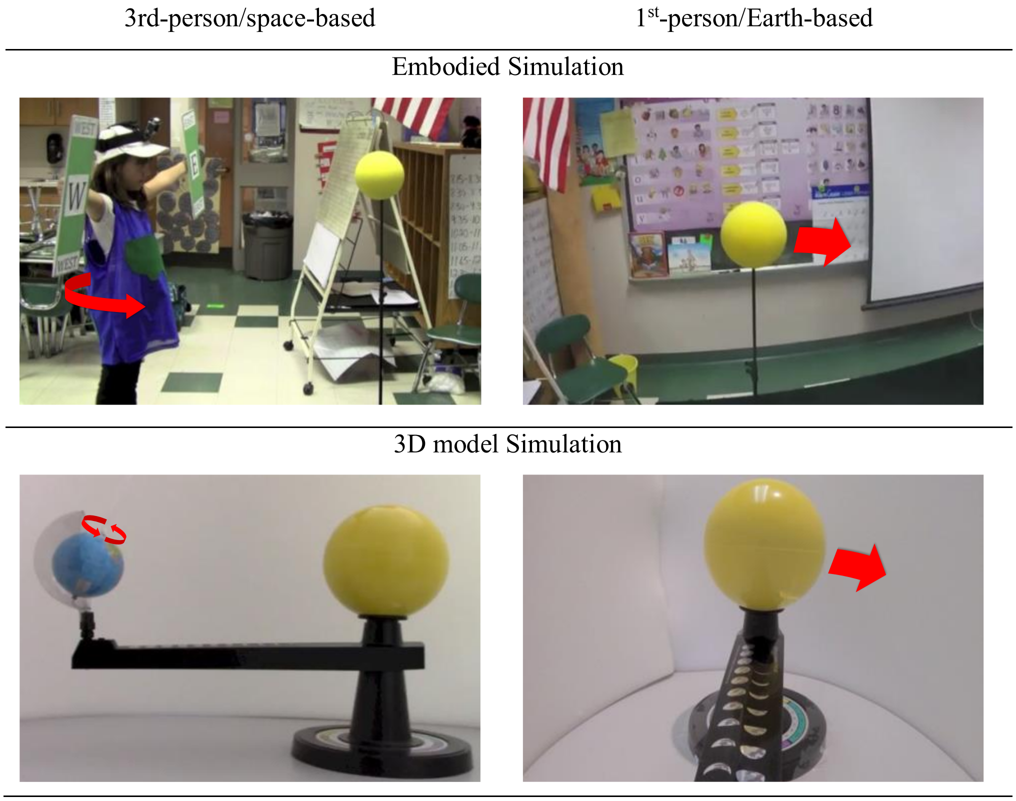
Verbal Ability. The Peabody Picture Vocabulary Test (PPVT), 4th edition (Dunn & Dunn, 2007) measures receptive vocabulary. Four pictures are presented on each trial. The test administrator says a word, and the participant must point to the picture to which it corresponds.

Mental Rotation. The spatial relations subtest of the Primary Mental Abilities Test (PMA-SR; Thurstone & Thurstone, 1962) measures mental rotation ability. On each trial, the participant must identify a rotated shape that forms a complete square with a second part.

Perspective Taking. The Perspective Taking Test for Children (PTT-C; Frick et al., 2014) involves identifying from a set of four options the picture taken by a toy photographer within a simple scene. The objects in the scene and the photographer's position vary across trials.

Instructional Activities

Embodied simulation. The activity followed a script based on the lesson plan for *Kinesthetic Astronomy* (Morrow & Zawaski, 2004). After orienting the participant to their role as Earth (see Figure 2, top left), the researcher guided them through a simulated 24-hour day. The participant's 1st-person observations were recorded using a head-mounted GoPro camera (see Figure 2, top right). A videorecorder on a tripod captured the session from a 3rd-person perspective. At the end of the activity, the participant performed one slow, careful rotation that supplied video footage for relational scaffolding.



3rd-person/space-based

1st-person/Earth-based

Embodied Simulation

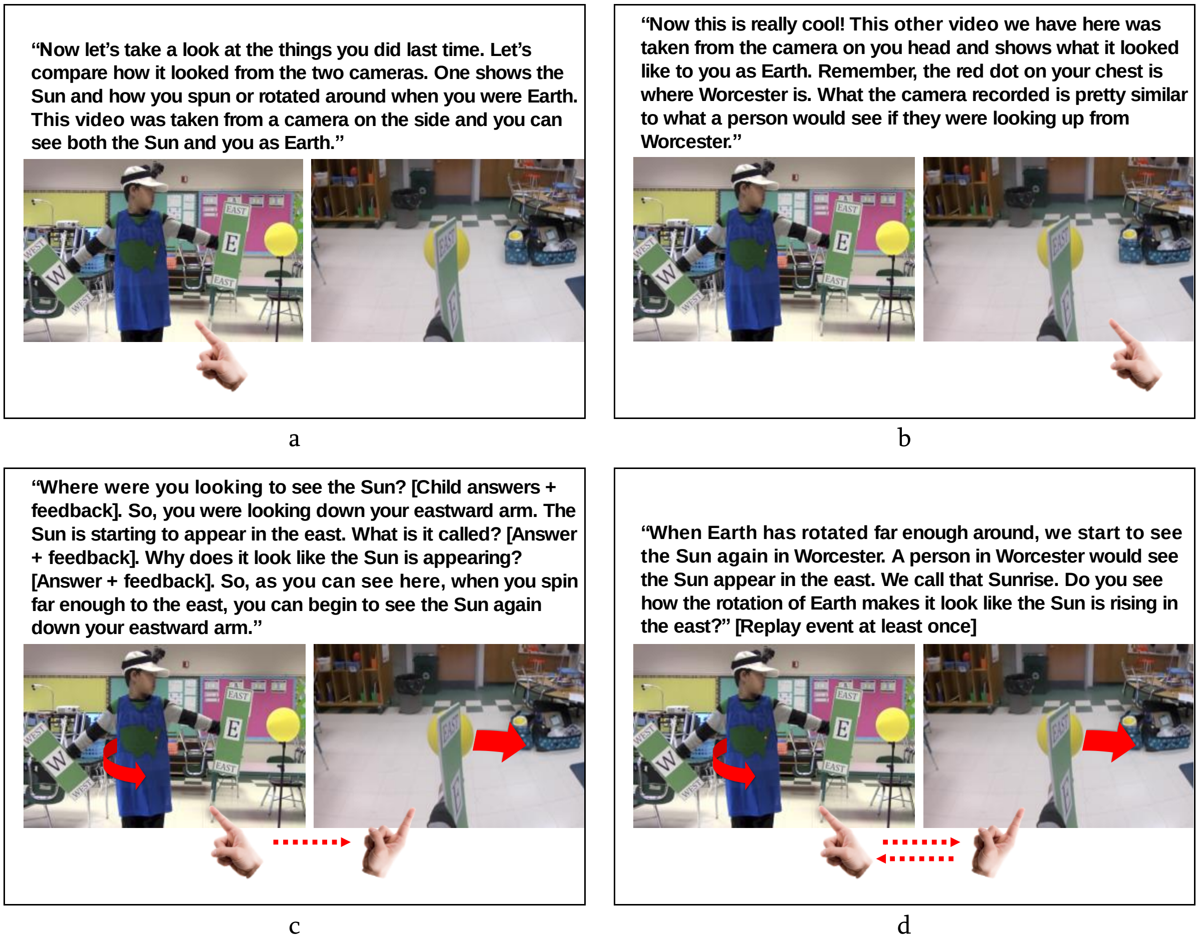
3D model Simulation

Figure 2. Screen captures of relational scaffolding video footage from an embodied simulation (top row) and 3D model simulation (bottom row). Note: Arrows indicate path of motion/apparent motion.

3D model simulation. This activity involved a physical 3D model with a Sun and rotating Earth (see Figure 2, bottom left). The simulation followed a detailed script similar in content and structure to the embodied simulation. The participant was prompted to look out from behind model Earth to adopt an Earth-based perspective for key events.

Relational Scaffolding 1. This activity used video of the embodied simulation. The participant's 3rd- and 1st-person videos were edited to create an approximately 20-30 second split-screen video of Earth's rotation as seen from each perspective. A trained researcher guided the participant through footage of midday, sunset, midnight, and sunrise, shown on a 13-inch MacBook Pro computer (see Relational Scaffolding 1 Script in the Supplemental Online

Materials). The researcher pointed at and between the videos to convey the correspondences, repeating key events several times (see Figure 3 for an excerpt).



“Now let’s take a look at the things you did last time. Let’s compare how it looked from the two cameras. One shows the Sun and how you spun or rotated around when you were Earth. This video was taken from a camera on the side and you can see both the Sun and you as Earth.”

a

“Now this is really cool! This other video we have here was taken from the camera on you head and shows what it looked like to you as Earth. Remember, the red dot on your chest is where Worcester is. What the camera recorded is pretty similar to what a person would see if they were looking up from Worcester.”

b

“Where were you looking to see the Sun? [Child answers + feedback]. So, you were looking down your eastward arm. The Sun is starting to appear in the east. What is it called? [Answer + feedback]. Why does it look like the Sun is appearing? [Answer + feedback]. So, as you can see here, when you spin far enough to the east, you can begin to see the Sun again down your eastward arm.”

c

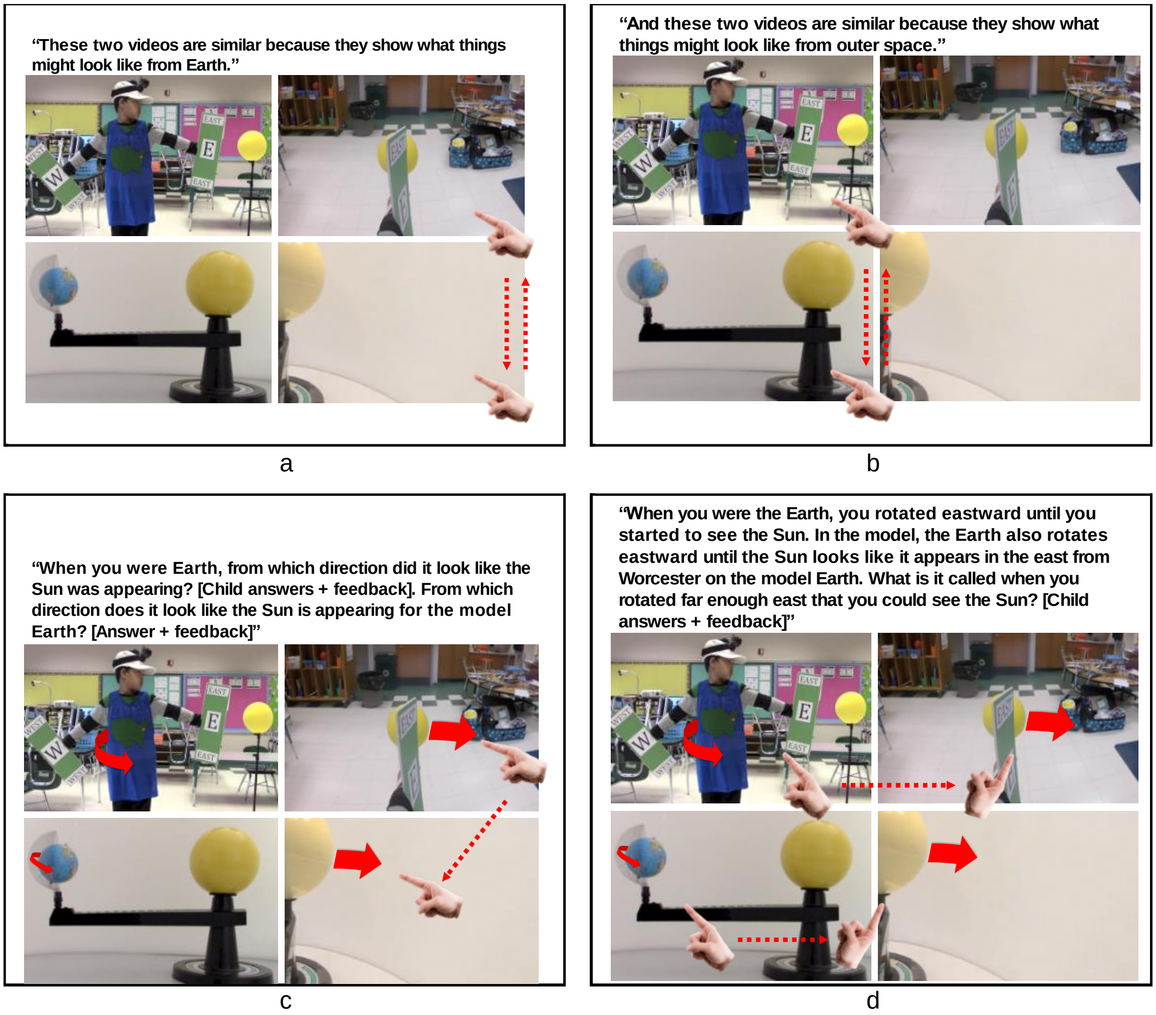
“When Earth has rotated far enough around, we start to see the Sun again in Worcester. A person in Worcester would see the Sun appear in the east. We call that Sunrise. Do you see how the rotation of Earth makes it look like the Sun is rising in the east?” [Replay event at least once]

d

Figure 3. Schematic Diagram of Relational Scaffolding 1. The excerpt is about *sunrise*. Each panel a-d shows a screenshot from video of 3rd-person/space-based perspective (on the left) and 1st-person/Earth-based perspective (on the right). Text from the session script is quoted above the images. Bold text signifies when the researcher pointed to one of the perspectives. Arrows were added to the images to indicate the participant’s motion (3rd-person) and the model Sun’s apparent motion (1st-person) as they appeared in the video. A single hand indicates a pointing gesture from the researcher. A single dotted arrow indicates a sequential gesture between the two videos. Two dotted arrows indicate back-and-forth gesturing between the videos. The format for this diagram is based on Yuan, Uttal, and Gentner (2017).

Relational Scaffolding 2. This activity used video from both the embodied and 3D model simulations. The footage was displayed on a 13-inch MacBook Pro computer in a 2 x 2 matrix, as in Figure 4. A trained researcher followed a script (Relational Scaffolding 2 in the

Supplemental Online Materials) that highlighted corresponding perspectives (see Figure 4 panels a and b), and the higher-order relations between the simulations (see Figure 4 panels c and d).



"These two videos are similar because they show what things might look like from Earth."

"And these two videos are similar because they show what things might look like from outer space."

"When you were Earth, from which direction did it look like the Sun was appearing? [Child answers + feedback]. From which direction does it look like the Sun is appearing for the model Earth? [Answer + feedback]"

"When you were the Earth, you rotated eastward until you started to see the Sun. In the model, the Earth also rotates eastward until the Sun looks like it appears in the east from Worcester on the model Earth. What is it called when you rotated far enough east that you could see the Sun? [Child answers + feedback]"

Figure 4. Schematic Diagram of Relational Scaffolding 2. The material is excerpted from the section of the session about *sunrise*. Text from the session script is quoted above the images. Bold text signifies when the researcher pointed to an event from one of the perspectives. The conventions from Figure 3 for motion arrows in the images and gesture arrows for the hands were used here also.

Procedure

Participants completed one study session per day, twice a week for three weeks at their school during after-school hours. The delayed posttest occurred about 6-7 weeks later.

Participants met with the same researcher in a quiet room (usually a classroom) each day. Each session lasted about 20-30 minutes. The sessions were videotaped (with parental permission).

In session 1, participants completed the vocabulary test and mental rotation test. Session 2 consisted of the day-night cycle knowledge interview (pretest), space science concept inventory items, and the perspective taking test. Sessions 3-5 varied between conditions. Participants in the RS condition completed the embodied simulation in session 3, relational scaffolding 1 in session 4, and both the 3D model simulation and relational scaffolding 2 in session 5. Participants in the No RS condition completed the embodied simulation in session 3 and again in session 4, followed by the 3D model simulation in session 5. Participants in the Models Only condition completed the 3D model simulation in sessions 3, 4, and 5. Participants in the Control condition did not complete instructional sessions. They remained in a supervised classroom during this time. In all conditions, session 6 consisted of the knowledge interview (posttest) and space science concept inventory items. Session 7 included the knowledge interview (delayed posttest), the mental rotation test, and the perspective taking test. Equivalent forms of the mental rotation test, perspective taking test, and science concept inventory items were administered in counterbalanced order across participants in each condition. Table 1 provides an overview of the procedure.

Table 1

Overview of procedure for Experiments 1 and 2

	Session						
	Pretest		Instruction			Posttest	Delayed Posttest
Condition	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
Experiment 1							
Control	Verbal ability test Mental rotation test	Day-night cycle interview Concept inventory items Perspective taking test	No instruction	No instruction	No instruction	Day-night cycle interview Concept inventory items	Day-night cycle interview Mental rotation test Perspective taking test
Models Only (MO)			3D model simulation	3D model simulation	3D model simulation		
No Relational Scaffolding (No RS)			Embodied simulation	Embodied simulation	3D model simulation		
Relational Scaffolding (RS)			Embodied simulation	Relational scaffolding 1	3D model simulation Relational scaffolding 2		
Experiment 2							
Sequential Scaffolding (SS)	Verbal ability test Mental rotation test	Day-night cycle interview Concept inventory items Perspective taking test	Embodied simulation	Sequential scaffolding 1	3D model simulation Sequential scaffolding 2	Day-night cycle interview Concept inventory items	Day-night cycle interview Mental rotation test Perspective taking test
Relational Scaffolding (RS)			Embodied simulation	Relational scaffolding 1	3D model simulation Relational scaffolding 2		

Coding the Day-night Cycle Interviews

We created a 27-component coding rubric, building on prior measures of space science understanding (e.g., Plummer et al., 2014; Vosniadou & Brewer, 1994). The components represent scientifically consistent ideas scored as correct/present vs. incorrect/absent based on the participant's interview responses. The internal consistency of the rubric was .80 (Cronbach's α) at pretest, .88 at posttest, and .86 at delayed posttest. The 27 components and information about the coding process are provided in the Supplemental Online Materials.

Results

Table 2 shows the results for the demographic, cognitive ability, and knowledge measures. There were no pre-instruction differences between conditions on any variable (gender: $\chi^2 = 2.58$, p =. 46; all other measures: Fs < 1.40, ps > .25).

Table 2

Demographic and Cognitive Variables for Experiment 1

Variable	Control *M* (*SD*)	Models Only *M* (*SD*)	No RS *M* (*SD*)	RS *M* (*SD*)	Overall *M* (*SD*)
			Demographics		
Gender (F:M)	14:13	18:9	13:14	17:10	62:46
Age	8.7 (.6)	8.6 (.5)	8.6 (.4)	8.7 (.5)	8.6 (.5)
Pre-/Posttest N	27	27	27	27	108
			Pre-instruction		
Verbal ability (PPVT Standard Score)	100.4 (13.8)	100.8 (16.3)	99.3 (13.5)	99.5 (13.9)	100.0 (14.2)
Pre mental rotation (PMA-SR)	9.8 (2.4)	9.2 (3.0)	8.8 (2.7)	8.3 (3.3)	9.0 (2.9)
Pre perspective taking (PTT-C)	10.2 (4.0)	8.8 (3.8)	9.1 (3.5)	8.2 (3.1)	9.1 (3.6)
Pre space science concept inventory (ASSCI)	2.7 (1.5)	2.7 (1.1)	2.9 (1.5)	3.0 (1.3)	2.8 (1.3)
Pre day-night understanding	8.5 (4.1)	7.7 (4.4)	7.3 (3.8)	7.2 (3.6)	7.7 (4.0)
			Post-instruction		
Post space science concept inventory (ASSCI)†	3.0 (1.7)	3.0 (1.2)	2.9 (1.2)	2.5 (1.3)	2.8 (1.4)
Post day-night understanding	7.8 (4.0)	13.3 (5.2)	12.3 (4.7)	15.8 (4.9)	12.3 (5.5)
			Delayed post-instruction		
Delay mental rotation (PMA-SR)‡	9.8 (2.8)	8.9 (2.4)	9.7 (2.5)	10.0 (3.3)	9.6 (2.8)
Delay perspective taking (PTT-C)§	12.0 (4.0)	10.0 (3.3)	11.3 (4.7)	9.9 (4.0)	10.8(4.1)

Delay day-night understanding	9.1 (4.8)	12.4 (4.9)	11.1 (5.0)	13.5 (4.6)	11.5 (5.0)
Delay interval (weeks)	7.1 (2.8)	6.3 (2.2)	6.8 (2.5)	6.6 (2.4)	6.7 (2.5)
Delayed Posttest N	19	19	24	21	83

Note. RS= Relational Scaffolding. †Post space science concept inventory , N = 98. ‡Delay mental rotation, N = 77. §Delay perspective taking, N = 81.

We conducted a multiple regression analysis to predict posttest day-night cycle understanding (score on the 27-component rubric) from the participant's *age*, *gender* (M=0, F=1), *pretest verbal*, *mental rotation*, and *perspective taking* scores, and *pretest day-night understanding*. Table 3 shows the zero-order correlations between variables. We also included a set of orthogonal contrasts in the regression model (Davis, 2010) to test: 1) the effect of receiving instruction vs. none at all: *instructional conditions* vs. *Control*, 2) the effect of receiving embodied and 3D model simulation vs. 3D model simulation alone: *No RS and RS* vs. *Models Only*, and 3) the effect of receiving relational scaffolding specifically: *No RS* vs. *RS*. Because the *No RS* vs. *RS* contrast directly tested the relational scaffolding effect, we crossed this factor with pretest understanding to test the hypothesized relational scaffolding x prior knowledge interaction (West, Aiken, & Krull, 1996). We included this interaction in the regression along with two others that crossed *No RS* vs. *RS* with pretest *mental rotation* and *perspective taking* scores. All predictors were mean-centered for the analysis.

Table 3

Pearson Correlations Between Experiment 1 Variables

Variable	2	3	4	5	6	7	8	9	10	11	12
1. Gender	-.01	.10	-.06	.10	-.02	.20*	-.07	.18	.05	-.05	.05
2. Age	-	-.43**	.02	-.08	-.05	-.16	-.07	-.13	.04	-.03	-.11
3. Verbal ability		-	.23*	.24*	.11	.33**	.29**	.27**	.21	.19	.27*

4. Pre mental rotation	-	.30**	.13	.14	.17	.06	.30**	.24*	.13
5. Pre perspective taking		-	.17	.38**	.12	.18	.21	.71**	.13
6. Pre space science concept inventory			-	.07	.32**	.21*	.11	.22*	.07
7. Pre day-night understanding				-	.18	.40**	.09	.32**	.46**
8. Post space science concept inventory†					-	.12	-.16	.23*	.05
9. Post day-night understanding						-	.21	.21	.64**
10. Delay mental rotation‡							-	.24*	.16
11. Delay perspective taking§								-	.17
12. Delay day-night understandingᵇ									-

Note. * $p < .05$, ** $p < .01$. For Gender: M=0, F=1. †Post space science concept inventory, N = 98. ‡Delay mental rotation, N = 77. §Delay perspective taking, N = 81. ᵇDelay day-night understanding, N = 83.

The regression model accounted for 53% of the variance in posttest understanding, $F(12,95) = 9.02$, $SEE = 3.98$, $p < .0001$. Figure 5 conveys the main findings of the orthogonal contrasts. Participants in the instructional conditions learned more than Control participants (β = .55, 95% CI (confidence interval) = [.40, .69], $p < .0001$, pr^2 (squared partial correlation) = .38). Participants who received embodied simulation (No RS and RS) learned about as much as Models Only participants (β = .09, 95% CI = [-.05, .23], $p = .20$, $pr^2 = .02$). Importantly, participants in the RS condition acquired the greatest understanding, significantly more than No RS participants (β = .24, 95% CI = [.09, .38], $p = .002$, $pr^2 = .10$). (We confirmed in a post hoc analysis—restructuring the orthogonal contrasts—that RS participants also achieved higher understanding than Models Only participants, β = .20, 95% CI = [.06, .34], $p = .009$, $pr^2 = .07$.) *Pretest day-night understanding* was also significant predictor in the model (β = .39, 95% CI =

[.22, .55], $p < .0001$, $pr^2 = .19$); however, no other factor or interaction was, including relational scaffolding x prior knowledge (βs = -.03 - .11, *p*s > .20).

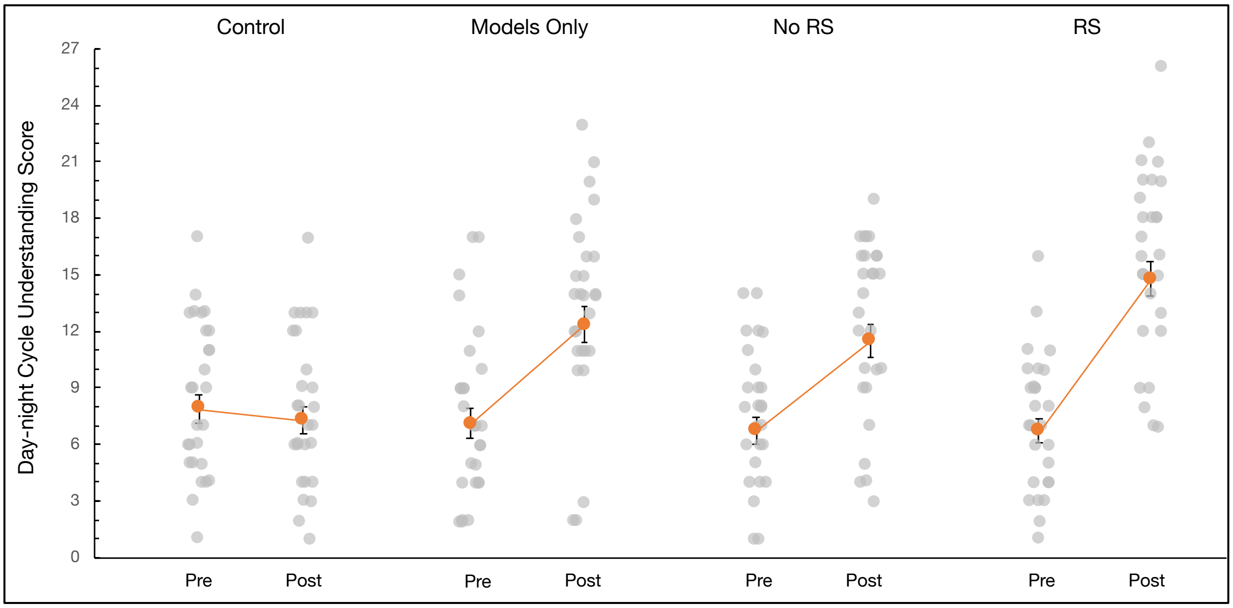


Figure 5. Pre- and posttest understanding scores for the day-night cycle knowledge interview. Individual participant scores shown in gray; condition means (± standard errors) in orange. Individual scores were jittered to produce separation on the x-axis.

Participants' broader space science knowledge changed little (see space science concept inventory means in Table 2). A 2 (test session) x 4 (condition) mixed ANOVA found no significant difference in mean scores on the concept inventory items from pre- to posttest, $F(1,94) < 1$, $p = .87$, $\eta_p^2 < .01$, nor was there a difference between conditions, $F(3,94) < 1$, $p = .97$, $\eta_p^2 < .01$ or a test session x condition interaction, $F(1,94) = 1.0$, $p = .32$, $\eta_p^2 = .03$.

Delayed Posttest

Eighty-three participants (77%) completed the delayed posttest. Day-night understanding scores were analyzed in a multiple regression analysis. To preserve statistical power, we included as predictors only those variables that were significant in the posttest analysis—*Pretest understanding,* the *instructional conditions* vs. *Control* contrast*,* and the *No RS* vs. *RS* contrast.

(An analysis that included all of the posttest predictors yielded the same pattern of results.) *Delay interval* (weeks between posttest and delayed posttest) was added as a predictor. All variables were mean-centered for the analysis.

The regression model accounted for 35% of the variance in delayed posttest understanding, $F(4,78) = 5.89$, $SEE = 10.64$, $p < .0001$. The results mirrored those at posttest. Participants in the instructional conditions maintained higher understanding than those in the Control condition ($\beta = .30$, 95% CI = [.12, .49], $p < .01$, $pr^2 = .12$). Participants in the RS condition had higher understanding than No RS participants (see Table 2); however, this difference diminished and was not statistically significant ($\beta = .18$, 95% CI = [-.01, .36], $p = .06$, $pr^2 = .05$). *Pretest understanding* was a significant predictor ($\beta = .52$, 95% CI = [.33, .71], $p < .0001$, $pr^2 = .28$). *Delay interval* was not significant ($\beta = -.13$, 95% CI = [-.32, .06], $p = .18$, $pr^2 = .02$), though understanding scores decreased with longer intervals.

We also analyzed participants' spatial test performance before vs. after instruction for both Experiments 1 and 2. The details are provided in the Supplemental Online Materials. In short, we found no change in mental rotation performance in either experiment. Perspective taking scores increased significantly in both experiments; however, the Control and instructional conditions showed about the same level of improvement.

Discussion

Participants in the instructional conditions greatly increased their understanding of the day-night cycle (though not of space science concepts more broadly). Participants who received relational scaffolding gained the most knowledge. These effects were somewhat attenuated by the delayed posttest (6-7 weeks after instruction), but the general pattern remained.

The results are consistent with our hypothesis that systematic comparison of observable and modeled events facilitates understanding of scientific models. Yet, further experimentation is required to separate the effects of comparison from other aspects of the RS condition that may enhance learning. Notably, only the RS condition included videos of observable and modeled events. Viewing this footage could reduce extraneous cognitive load (Sweller, 1994), enabling a student to devote limited mental resources to sense-making processes (Mayer & Moreno, 2003). It is precisely for topics that require attention to relations within a system (high *element interactivity*) that a reduction in extraneous load should be especially beneficial (Sweller, 1994).

Experiment 2 teased apart guided comparison from the viewing of video footage. We compared the RS condition to a condition in which videos of observable and modeled events were presented *sequentially*, without explicit comparison. If guided comparison is integral to the relational scaffolding effect, then the RS condition will produce greater understanding than the sequential video condition. Zeroing in on the role of guided comparison also permits a more direct test of the hypothesis that comprehensive, guided comparisons are especially helpful for students with little prior understanding. If so, any advantage of the RS condition should be most pronounced among lower-knowledge students.

In Experiment 2 a new group of 3rd-grade participants received day-night cycle instruction supplemented with either relational scaffolding or *sequential* scaffolding (SS), in which videos of observable (Earth-based) and modeled (space-based) events were presented one after the other. Experiment 2 also varied whether the scaffolding involved video of the participant's own embodied simulation (*self* footage) vs. research assistant's (*stock* footage). If relational scaffolding requires personalized footage, it would be challenging to implement in formal educational settings. The footage variable was crossed with the scaffolding manipulation to create four experimental conditions.

Experiment 2

Method

Participants. One hundred twenty-eight 3rd-grade children were sampled from elementary schools in Worcester, Massachusetts. Twenty-eight (22%) dropped out before the posttest. One was removed from the dataset because their PPVT score was more than 2 *SD* below age level. Our remaining sample was 99 3rd-graders (59 females, 40 males; $M_{age} = 8.6$ years, *SD* = 0.4). This sample size provides > .90 power to detect a medium-to-large effect (Cohen's $f^2 \geq .25$) for our planned analysis (Faul et al., 2009). Participants were randomly assigned to one of the four conditions (25 per condition; the SS-Self Footage condition had 24).

Materials. Experiment 2 used the measures and instructional activities from Experiment 1, and two new sequential scaffolding activities. ***Sequential scaffolding 1*** used footage of the embodied simulation, either *self* or *stock*. The 1st-person perspective was shown first. Although not shown simultaneously, the participant was prompted to attribute the Sun's apparent motion to Earth's rotation (3rd-person perspective). The 3rd-person perspective was shown second, and likewise referenced the 1st-person perspective. ***Sequential scaffolding 2*** presented video of the 3D model simulation, first from an Earth-based perspective and then from a space-based perspective. The activity used a similar script to sequential scaffolding 1. When stock footage was shown, the script adopted 3rd-person reference (e.g., "Remember what *they* recorded from the camera on *their* head…").

Procedure

Participants completed the procedure within their school during after-school hours. They completed one session per day, twice a week for three weeks. The delayed posttest was about four weeks later (sooner than in Experiment 1 to accommodate a school break). Sessions 1 and 2 (pretest), 6 (posttest), and 7 (delayed posttest) were the same as in Experiment 1. The instructional sessions (sessions 3-5) for the Relational Scaffolding condition were equivalent to the RS condition from Experiment 1. Participants in the Sequential Scaffolding (SS) condition completed the embodied simulation in session 3, sequential scaffolding 1 in session 4, and both the 3D model simulation and sequential scaffolding 2 in session 5. Table 1 provides an overview of the procedure.

Coding the Interviews

We used the 27-component rubric from Experiment 1. The internal consistency of the instrument was .81 (Cronbach's α) at pretest, .85 at posttest, and .83 at delayed posttest.

Results

Table 4 shows the results for the demographic and other variables. There were no pre-instruction differences between conditions in gender composition ($\chi^2 = 0.89$, p =. 35) or any other variable, ts < 1.25, ps > .20.

Table 4

Demographic and Cognitive Variables for Experiment 2

	Sequential Scaffolding		Relational Scaffolding		Overall
Variable	Stock *M* (*SD*)	Self *M* (*SD*)	Stock *M* (*SD*)	Self *M* (*SD*)	*M* (*SD*)
			Demographics		
Gender (F:M)	12:13	20:4	12:12	14:11	59:40
Age	8.6 (0.4)	8.5 (0.4)	8.6 (0.5)	8.7 (0.4)	8.6 (0.4)
Pre-/Posttest N	25	24	25	25	99
			Pre-instruction		

Verbal ability (PPVT Standard Score)	103.0 (13.7)	103.9 (15.3)	100.0 (12.9)	100.0 (14.5)	101.7 (14.0)
Pre mental rotation (PMA-SR)	9.0 (2.2)	9.3 (1.9)	9.4 (2.2)	9.5 (2.4)	9.3 (2.2)
Pre perspective taking (PTT-C)	9.7 (2.8)	10.0 (3.6)	10.8 (4.1)	10.5 (3.7)	10.3 (3.6)
Pre space science concept inventory (ASSCI)	2.7 (1.4)	2.5 (1.4)	2.4 (1.7)	2.8 (1.2)	2.6 (1.4)
Pre day-night understanding	6.8 (3.6)	8.1 (4.3)	7.4 (5.2)	6.6 (4.2)	7.2 (4.3)
Post-instruction					
Post space science concept inventory (ASSCI)[†]	3.3 (1.5)	2.9 (1.5)	3.3 (1.4)	3.0 (1.7)	3.1 (1.5)
Post day-night understanding	13.6 (5.9)	14.5 (6.0)	15.2 (4.7)	13.8 (5.2)	14.3 (5.4)
Delayed post-instruction					
Delay mental rotation (PMA-SR)[‡]	9.3 (2.6)	9.9 (2.8)	9.5 (3.1)	9.4 (2.1)	9.5 (2.6)
Delay perspective taking (PTT-C)[§]	11.6 (4.1)	11.5 (4.9)	13.0 (3.8)	13.0 (4.2)	12.3 (4.2)
Delay day-night understanding	12.7 (5.5)	13.6 (4.6)	15.3 (4.6)	16.2 (5.6)	14.3 (5.1)
Delay interval (weeks)	3.9 (1.6)	3.9 (1.8)	3.7 (1.3)	4.5 (1.9)	4.0 (1.6)
Delayed Posttest N	19	16	22	13	70

Note. [†]Post space science concept inventory, N = 94. [‡]Delay mental rotation, N = 51. [§]Delay perspective taking, N = 68.

We conducted a multiple regression analysis to predict posttest understanding from *age*, *gender* (M=0, F=1), pretest *verbal*, *mental rotation*, *perspective taking* scores, *pretest understanding, video condition* (stock=0, self=1), and *comparison condition* (sequential scaffolding=0, relational scaffolding=1). In this model, *comparison condition* directly tests the effect of guided comparison. We crossed this factor with pretest understanding to test the hypothesized guided comparison x prior knowledge interaction. We also crossed with comparison condition with pretest *mental rotation* and *perspective taking* scores to explore possible interactions with spatial ability. Predictors were mean-centered for the analysis. Table 5 shows the zero-order correlations between variables.

Table 5

Pearson Correlations Between Experiment 2 Variables

Variable	2	3	4	5	6	7	8	9	10	11	12
1. Gender	.08	-.05	.11	<.01	.06	.09	.11	.02	.03	-.04	-.06
2. Age	-	-.38**	-.01	-.09	-.05	.01	-.10	<.01	.19	.01	.12
3. Verbal ability		-	.27**	.25*	.19	.23*	.26*	.22*	.01	.09	.01
4. Pre mental rotation			-	.05	.11	.37**	.27**	.23*	.32*	.16	.01
5. Pre perspective taking				-	.03	.29**	.24*	.21*	.02	.53**	.13
6. Pre space science concept inventory					-	.13	.46**	.11	.18	.06	.22
7. Pre day-night understanding						-	.36**	.46**	.21	.25*	.33**
8. Post space science concept inventory†							-	.42**	.13	.23	.37**
9. Post day-night understanding								-	.12	.34**	.63**
10. Delay mental rotation‡									-	.26	.02
11. Delay perspective taking§										-	.19
12. Delay day-night understanding^{þ}											-

Note. $^{*}p < .05$, $^{**}p < .01$. For Gender, F=1, M=0. †Post space science concept inventory, N = 94. ‡Delay mental rotation, N = 51. §Delayed perspective taking, N = 68. ^{þ}Delay day-night understanding, N = 70.

The regression model accounted for 31% of the variance in posttest day-night cycle understanding, $F(11,87) = 3.52$, $SEE = 4.79$, $p < .001$. The main finding was a significant *comparison condition* x *pretest understanding* interaction ($\beta = -.41$, 95% CI = [-.73, -.09], $p = .01$, $pr^2 = .07$). Figure 6 (left panel) shows mean posttest day-night cycle understanding scores for RS and SS participants who scored below and above the median on the pretest (i.e., low vs.

high initial knowledge). For low-knowledge (but not high-knowledge) participants, there was a clear advantage of RS over SS. Neither manipulated variable—*video condition* (β = -.07) nor *comparison condition* (β = .07)—was itself a significant predictor, *ps* > .45. *Pretest understanding* was significant (β = .72, 95% CI = [.40, 1.0], $p < .0001$, $pr^2 = .19$); however, no other variable was (βs =.01 - .12, *ps* > .25), nor was there another interaction (βs = .01, *ps* = .96).

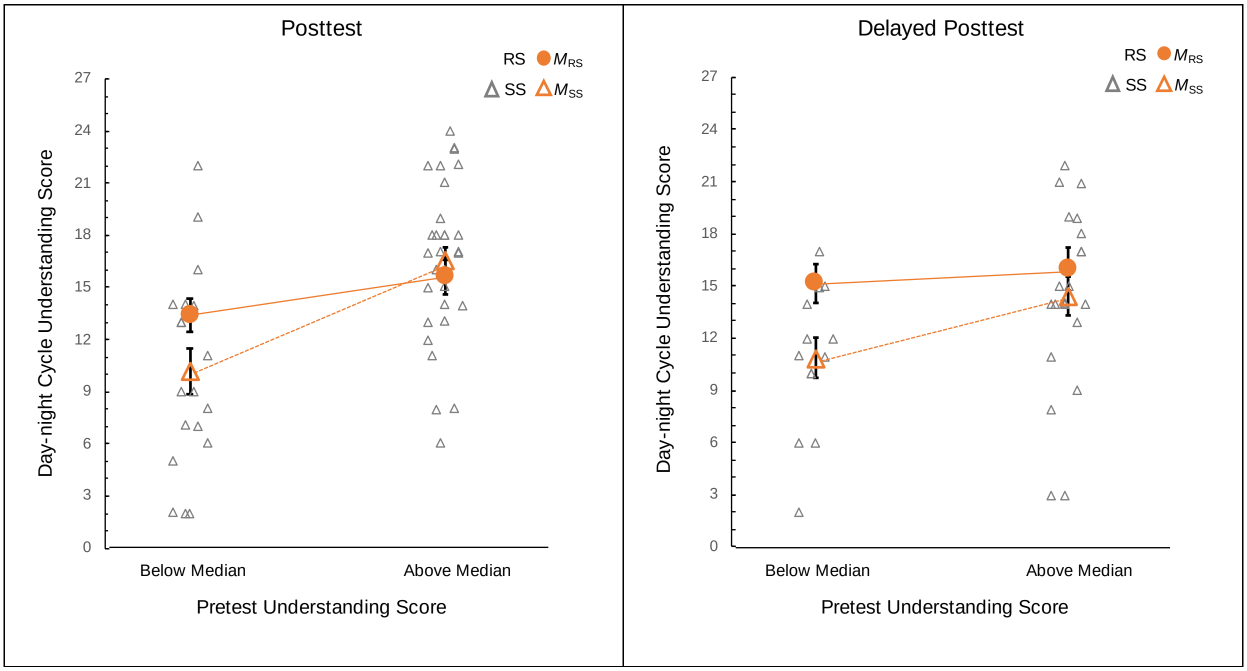


Figure 6. Posttest (left panel) and delayed posttest day-night cycle understanding scores (right panel) for participants who scored below and above the median on the pretest. Individual participant scores shown in gray; means (± standard errors) in orange. Individual scores were jittered to produce separation on the x-axis. RS = Relational Scaffolding, SS = Sequential Scaffolding.

Participants' scores on the space science concept inventory (ASSCI) increased slightly, by about 0.5 items (see Table 3). A 2 (test session) x 2 (comparison condition) x 2 (video condition) mixed ANOVA revealed this increase was significant, $F(1,90) = 11.96$, $p < .001$, η_p^2 = .12, but there was no effect of comparison condition or video condition, *F*s < 1, *ps* > .60, η_p^2s < .01, and no interactions, *F*s < 1.95, *ps* > .15, η_p^2s < .03.

Delayed Posttest

Seventy participants (71%) completed the delayed posttest. An analysis of spatial test performance before vs. after instruction again appears in the Supplemental Online Materials. We analyzed delayed day-night understanding in a multiple regression analysis. To preserve power, we included as predictors only those variables involved in the significant *pretest understanding* x *comparison condition* interaction from posttest: *pretest understanding, comparison condition,* and the *interaction* term. We also added *delay interval* (in weeks) as a predictor. All variables were mean-centered for the analysis.

The regression model accounted for 23% of the variance in delayed posttest understanding, $F(4,65) = 4.94$, $SEE = 4.61$, $p < .01$. The *pretest understanding* x *comparison condition* interaction was not significant ($\beta = -.32$, 95% CI = [-.67, .03], $p = .07$, $pr^2 = .05$). Instead, *comparison condition* emerged as a significant predictor ($\beta = .27$, 95% CI = [.06, .49], $p < .05$, $pr^2 = .09$). Figure 6 (right panel) shows that mean delayed posttest understanding was overall higher in the RS condition, although this RS advantage was especially pronounced among low-knowledge participants. *Pretest understanding* was also a significant predictor ($\beta = .60$, 95% CI = [.24, .95], $p < .01$, $pr^2 = .15$), but *delay interval* was not ($\beta = -.13$, 95% CI = [-.35, .09], $p = .25$).

Discussion

Experiment 2 found that participants with relatively low initial knowledge benefitted from explicit, guided comparisons between videos of Earth- and space-based perspectives—the RS condition. Viewing the same videos without comparison—the SS condition—was inferior. This RS advantage emerged across levels of prior knowledge by the delayed posttest, though lower knowledge participants remained the primary beneficiaries.

The second manipulated variable—whether scaffolding involved the participant's personal observations (*self* footage) vs. another person's (*stock* footage)—was unrelated to learning outcomes. We note that the stock footage showed the same materials, setup, and sequence as the participant's own embodied simulation. It is possible that stock footage will be less effective if these properties are altered. That said, the results are encouraging for the prospect of implementing relational scaffolding in educational contexts, like classrooms, where individualized footage is unfeasible.

General Discussion

Our findings demonstrate that relational scaffolding—systematic, guided comparison of observable and corresponding modeled events—supports students' understanding of scientific models. We tested relational scaffolding in the context of instruction about the day-night cycle, a fundamental science topic that involves linking Earth-based observations to a space-based model of planetary motion. Across two experiments, relational scaffolding was found to enhance 3rd graders' understanding of the day-night cycle, especially students with little prior knowledge.

This research speaks to the potential of extending theories of analogical thinking to fundamental issues in science education (see also Goldwater & Schalk, 2016; Jee et al., 2010). When a scientific model has no obvious connection to observable phenomena, structural alignment may be crucial for comprehension. Relational scaffolding facilitates the alignment process through several coordinated supports. Explicit comparisons illuminated shared relational structure (Goldwater & Gentner, 2015). A trained researcher pointed at and between videos of observed and modeled events to clarify key correspondences (Richland et al., 2007; Yuan et al., 2017). The scaffolding was comprehensive, underscoring the coherence of the scientific model and discouraging inaccurate, piecemeal explanations (Au & Romo, 1996; Chi et al., 2012). This systematic approach may be most effective when models are complex, unfamiliar, or counterintuitive. Future applications of relational scaffolding should therefore consider the subject matter, students' background knowledge, and common conceptions that might impede (or promote) science learning (Shtulman, 2017; Vosniadou & Skopeliti, 2014).

Relational scaffolding is well suited to novices, who lack conceptual knowledge for inferring relational structure (McNamara et al., 1996) and are susceptible to cognitive overload during instruction (Sweller, 1994). An intriguing possibility is that relational scaffolding can help "level the playing field" for children who are poorly prepared for science education. Science achievement gaps emerge in the early school years and often persist, largely because of disparities in students' basic science knowledge (Morgan, Farkas, Hillemeier, & Maczuga, 2016). Methods that assist underprepared students would profoundly impact many children, and, ultimately, increase the pool of qualified candidates for careers in science.

Important questions remain about the model-based instruction that accompanies relational scaffolding. Our Experiment 1 found embodied simulation did not increase student learning beyond models-only instruction. This seems at odds with evidence that physical experience improves science concept learning (e.g., Kontra et al. 2015). However, rather than passively observing, participants repeatedly adopted an Earth-based perspective (looking out from model Earth) during the 3D model activity. A fully enacted simulation may be unnecessary when physical (or virtual) modeling is immersive, providing sensorimotor feedback linking model-based observations to embodied actions (DeSutter & Stieff, 2017; Lindgren, Tscholl, Wang, & Johnson, 2016). Nonetheless, our findings leave open the possibility that embodied simulation helps students understand an external model when the two are explicitly aligned—that is, through relational scaffolding. Further experimentation is required to explore this and other questions about how model-based instruction contributes to the relational scaffolding effect.

The day-night cycle was a prime candidate for testing relational scaffolding. Yet, many other fundamental scientific models portray invisible entities and processes with no obvious connection to observable phenomena. Seasonal change relates to Earth's tilt and orbit; energy transfer and state changes involve invisible molecular activity; and illness and immunity relate to microscopic viruses, vaccines, and immune cells. Nonscientific conceptions are widespread for each of these topics (Shtulman, 2017). Each is thus a promising subject matter for future tests of the relational scaffolding approach.

Author Contributions

Both authors created the study concept, developed the study design and materials, and supervised the training of research assistants. B. Jee developed the data coding, and performed the data analysis and interpretation. F. Anggoro recruited the participants and managed the data collection. B. Jee drafted the manuscript with F. Anggoro providing critical revisions. Both authors approved the final version of the manuscript for submission.

References

American Association for the Advancement of Science (2009). *Benchmarks for science literacy*. New York, NY: Oxford University Press.

Au T. K., Romo L. F., (1996). Building a coherent conception of HIV transmission. *The Psychology of Learning and Motivation, 35,* 193–241

Chi, M. T., Roscoe, R. D., Slotta, J. D., Roy, M., & Chase, C. C. (2012). Misconceived causal explanations for emergent processes. *Cognitive Science*, *36*(1), 1-61.

Clement, J. (1993). Using bridging analogies and anchoring intuitions to deal with students' preconceptions in physics. *Journal of Research in Science Teaching, 30*(10), 1241-1257.

DeSutter, D., & Stieff, M. (2017). Teaching students to think spatially through embodied actions: Design principles for learning environments in science, technology, engineering, and mathematics. *Cognitive research: principles and implications*, *2*(1), 22.

Davis, M. J. (2010). Contrast coding in multiple regression analysis: Strengths, weaknesses, and utility of popular coding structures. *Journal of Data Science*, *8*(1), 61-73.

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody picture vocabulary test*. Pearson Assessment.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149-1160.

Frick, A., Möhring, W., & Newcombe, N. S. (2014) Picturing perspectives: Development of perspective-taking abilities in 4- to 8-year-olds. *Frontiers in Psychology, 5*(386), 1-7.

Gentner, D., Anggoro, F., & Klibanoff, R. (2011). Structure-mapping and relational language support children's learning of relational categories. *Child Development, 82*(4) 1173-1188.

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist, 52*(1), 45-56.

Gentner, D., & Toupin, C. (1986). Systematicity and surface similarity in the development of analogy. *Cognitive Science*, *10*(3), 277-300.

Goldwater, M. B., & Gentner, D. (2015). On the acquisition of abstract knowledge: Structural alignment and explication in learning causal system categories. *Cognition*, *137*, 137-153.

Goldwater, M. B., & Schalk, L. (2016). Relational categories as a bridge between cognitive and educational research. *Psychological Bulletin*, *142*(7), 729-757.

Holyoak, K. J., & Koh, K. (1987). Surface and structural similarity in analogical transfer. *Memory and Cognition, 15*(4), 332-340

Jee, B. D., Uttal, D. H., Gentner, D., Manduca, C., Shipley, T. F., Tikoff, B., Ormand, C. J., & Sageman, B. (2010). Commentary: Analogical thinking in geoscience education. *Journal of Geoscience Education, 58*(1), 2-13.

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, *19*(4), 509-539.

Kontra, C., Lyons, D. J., Fischer, S. M., & Beilock, S. L. (2015). Physical experience enhances science learning. *Psychological science*, *26*(6), 737-749.

Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.

Kurtz, K. J., Miao, C. H., & Gentner, D. (2001). Learning by analogical bootstrapping. *The Journal of the Learning Sciences*, *10*(4), 417-446.

Lindgren, R., Tscholl, M., Wang, S., & Johnson, E. (2016). Enhancing learning and engagement through embodied interaction within a mixed reality simulation. *Computers & Education*, *95*, 174-187.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, *38*(1), 43-52.

McCloskey, M. (1983). Naive theories of motion. In D. Gentner, & A. L. Stevens (Eds.), *Mental models* (pp. 299-324). Hillsdale, NJ: Erlbaum.

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, *14*(1), 1-43.

Morgan, P. L., Farkas, G., Hillemeier, M. M., & Maczuga, S. (2016). Science achievement gaps begin very early, persist, and are largely explained by modifiable factors. *Educational Researcher, 45*(1), 18-35.

Morrow, C. A., & Zawaski, M. (2004). Kinesthetic Astronomy: Significant upgrades to the Sky Time lesson that support student learning. *Bulletin of the American Astronomical Society, 36*, 1562.

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Committee on a Conceptual Framework for New K-12 Science Education Standards. Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Nersessian, N. J. (2010). *Creating scientific concepts*. Cambridge, MA: MIT press.

Plummer, J. D., Kocareli, A., & Slagle, C. (2014). Learning to explain astronomy across moving frames of reference: Exploring the role of classroom and planetarium-based instructional contexts. *International Journal of Science Education, 36*(7), 1083-1106.

Richland, L. E., Zur, O., & Holyoak, K. J. (2007). Cognitive supports for analogies in the mathematics classroom. *Science*, *316*(5828), 1128-1129.

Sadler, P. M., Coyle, H., Miller, J. L., Cook-Smith, N., Dussault, M., & Gould, R. R. (2010). The astronomy and space science concept inventory: development and validation of assessment instruments aligned with the k–12 national science standards. *Astronomy Education Review, 8*(1), electronic ID: 010111.

Shtulman, A. (2017). *Scienceblind: Why our intuitive theories about the world are so often wrong.* New York, NY: Basic Books.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction*, *4*(4), 295-312.

Thurstone, T. G., & Thurstone, L. L. (1962). *Primary mental abilities tests*. Science Research Associates.

Vosniadou, S., & Brewer, W. F. (1994). Mental models of the day/night cycle. *Cognitive Science*, *18*(1) 123-183.

Vosniadou, S., & Skopeliti, I. (2014). Conceptual change from the framework theory side of the fence. *Science & Education*, *23*(7), 1427-1445.

Vosniadou, S., & Skopeliti, I. (2017). Is it the Earth that turns or the Sun that goes behind the mountains? Students' misconceptions about the day/night cycle after reading a science text. *International Journal of Science Education, 39*(15), 2027-2051.

West, S. G., Aiken, L. S., & Krull, J. L. (1996). Experimental personality designs: Analyzing categorical by continuous variable interactions. *Journal of Personality*, *64*(1), 1-48.

Yuan, L., Uttal, D., & Gentner, D. (2017). Analogical processes in children's understanding of spatial representations. *Developmental Psychology*, *53*(6), 1098-1114.

Supplementary Online Materials

Supplementary Methods

Items from the Astronomy and Space Science Concept Inventory (ASSCI). The ASSCI was developed to measure students' mastery of the astronomical concepts found in AAAS benchmarks and NRC standards for different grade levels (Sadler et al., 2010). This multiple-choice instrument contains items with a combination of difficulty, discrimination, and distractor popularity (based on common, but scientifically inaccurate responses). The full instrument is available for research and educational purposes upon request to the test developers. We selected 9 items that were broadly related to—but not covered in—our instructional sessions, such as, "What is the largest source of heat for the surface of Earth?" and "At what time of night should you try to see the North Star?" There were 7 items from the K-4 form and 2 from the 5-8 grade level instrument. A large-scale norming study found that students answered about 4.5 out of these 9 items correctly (Sadler et al., 2010).

Supplementary Results

Participant Attrition. The 36 participants who left of Experiment 1 before the posttest were about the same age, on average, as the rest of the sample (*M* = 8.5 years, *SD* = 0.5). Most attrition occurred before the instructional sessions began—the non-finishers completed a mean of 2.5 sessions (*Mode* = 2). Attrition was about evenly distributed across the 4 conditions. The 28 participants who left of Experiment 2 before the posttest were of similar age to the rest of the sample (*M* = 8.4 years, *SD* = 0.3). The non-finishers completed a mean of 2.6 sessions (*Mode* = 2). Attrition was about evenly distributed across the conditions. The reasons for attrition are unknown, but we suspect that the biggest contributor was the logistical strain on parents and caregivers who had to accommodate their child’s involvement in the weeks-long after-school program.

Coding the Day-night Cycle Knowledge Interviews. Supplemental Table 1 contains the full list of 27 components from the rubric. We established interrater reliability through independent coding trials followed by reliability analyses and discussion and refinement of the coding criteria. Coders were blind to the participant’s condition and test session (pre-, post-, or delayed posttest), and never coded a participant whom they had run through the study. Our criterion was reliability of .80 or higher (Krippendorff’s α) between each rater and the other coders on two consecutive rounds of 4-6 interviews. All research team members met this target level of reliability before contributing their coding to the data set. Intercoder reliability ranged from .82 - .96 for the Experiment 1 interviews and .83 - .92 for the Experiment 2 interviews.

Supplemental Table 1

Interview Coding Rubric Components

1. Sun is in sky during day (must refer to Sun)
2. Sun is not in sky during night (or describe sky as dark, see stars, moon sometimes, etc.)
3. Place on Earth faces Sun in day
4. Sun at other side of Earth, Place on Earth faces away from Sun at night
5. Earth spins/turns/rotates
6. Sun does not move
7. Rotation causes places to face Sun at different times
8. Our day is nighttime for the other side of the Earth
9. Our night is daytime for the other side of the Earth
10. Illustrate Sun path: Correct order and location of morning, noon, evening Sun
11. Illustrate Sun path: Sun moves from east to west in a day
12. Location on Earth rotates toward the Sun to cause “sunrise”
13. Earth rotates eastward to cause “sunrise”
14. Location on Earth rotates away from the Sun to cause “sunset”
15. Earth rotates eastward to cause “sunset”
16. Modeling how Earth moves in a day 1: Earth rotates (on any axis)
17. Modeling how Earth moves in a day 2: Sun does not move
18. Modeling how Earth moves in a day 3: Earth makes one complete rotation (no more)
19. Modeling day-night cycle 1: Dot (location of Worcester, MA) faces the Sun in the day

20. Modeling day-night cycle 2: Dot faces away from the Sun at night
21. Modeling day-night cycle 3: Earth rotates east from day to night
22. Modeling sunrise 1: Dot begins facing away from the Sun and turns eastward
23. Modeling sunrise 2: Dot turns eastward and begins to face the Sun
24. Modeling sunset 1: Dot begins facing the Sun and turns eastward
25. Modeling sunset 2: Dot turns eastward and begins to face away from the Sun
26. Link model to drawing 1: Earth rotates eastward
27. Link model to drawing 2: Eastward rotation makes Sun appear to move east-west

Note. Components 16-27 are based on the 3D modeling portion of the interview.

Spatial Ability Before and After Instruction

Experiment 1. Of the 108 participants, 77 completed the mental rotation test and 81 completed the perspective taking test at delayed posttest. We performed separate 2 (test session) x 4 (condition) mixed ANOVAs for mental rotation (PMA-SR) and perspective taking (PTT-C) test performance. Average mental rotation scores did not increase significantly from pre- to delayed posttest, $F(1,73) = 0.9$, $p = .36$, $\eta_p^2 = .01$, nor was there an effect of condition, $F(3,73) = 0.08$, $p = .97$, $\eta_p^2 < .01$, or test session x condition interaction, $F(3,73) = 1.2$, $p = .31$, $\eta_p^2 = .05$. Mean perspective taking scores, however, did increase significantly from pre- to delayed posttest, $F(1,77) = 29.3$, $p < .0001$, $\eta_p^2 = 0.28$. However, there was no effect of condition (see means in Table 2), $F(3,77) = 1.5$, $p = .22$, $\eta_p^2 = 0.06$, and no test session x condition interaction, $F(3,77) = .25$, $p = .86$, $\eta_p^2 = 0.01$. Because the Control (No Instruction) group showed about the same level of improvement as the instructional conditions, the overall improvement in perspective taking performance cannot be attributed to the emphasis on perspective taking in the instructional conditions.

Experiment 2. Of the 99 participants, 51 completed the mental rotation test and 68 completed the perspective taking test at delayed posttest. We performed separate 2 (test session) x 2 (comparison condition) x 2 (video condition) mixed ANOVAs for mental rotation and perspective taking test performance. Average mental rotation scores did not increase significantly from pre- to delayed posttest, $F(1,47) = 1.2$, $p = .27$, $\eta_p^2 = .03$, nor was there a main effect of comparison condition or video condition, *F*s < 1, *p*s > .50, or any interactions, *F*s < 1, *p*s > .35. Mean perspective taking scores increased significantly from pre- to delayed posttest (see means in Table 4), $F(1,64) = 12.0$, $p < .001$, $\eta_p^2 = 0.16$, but there was no effect of comparison condition or video condition, *F*s < 1, *p*s > .30, or any interactions, *F*s < 1.7, *p*s > .20. The overall improvement in perspective taking scores was similar in magnitude to the Experiment 1 effect—including the control condition. Thus, although perspective taking test performance generally improved from pre- to delayed posttest in each experiment, about the same amount of improvement occurred regardless of whether participants received instruction about the day-night cycle.